TECHNICAL RESEARCH REPORT

Foundations of Swarm Intelligence: From Principles to Practice

by Mark Fleischer

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **2003**	2. REPORT TYPE	3. DATES COVERED **-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Foundations of Swarm Intelligence: From Principles to Practice	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Army Research Laboratory,2800 Powder Mill Road,Adelphi,MD,20783**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES **14**	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**			

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Foundations of Swarm Intelligence:
From Principles to Practice

Mark Fleischer
Institute for Systems Research
University of Maryland
College Park, Maryland 20742

Abstract— Swarm Intelligence (SI) is a relatively new paradigm being applied in a host of research settings to improve the management and control of large numbers of interacting entities such as communication, computer and sensor networks, satellite constellations and more. Attempts to take advantage of this paradigm and mimic the behavior of insect swarms however often lead to many different implementations of SI. The rather vague notions of what constitutes self-organized behavior lead to rather *ad hoc* approaches that make it difficult to ascertain just what SI is, assess its true potential and more fully take advantage of it. This article provides a set of general principles for SI research and development. A precise definition of *self-organized behavior* is described and provides the basis for a more axiomatic and logical approach to research and development as opposed to the more prevalent *ad hoc* approach in using SI concepts.

The concept of *Pareto optimality* is utilized to capture the notions of efficiency and adaptability. A new concept, *Scale Invariant Pareto Optimality* is described and entails symmetry relationships and scale invariance where Pareto optimality is preserved under changes in system states. This provides a mathematical way to describe *efficient tradeoffs of efficiency* between different scales and further, mathematically captures the notion of the *graceful degradation of performance* so often sought in complex systems.

Index Terms— swarm intelligence, self-organization, multiobjective optimization, Pareto optima, finite-state machines

I. INTRODUCTION

TODAY'S communications networks have become enormously complex systems. New technologies from sensor networks, web-enabled PDAs, remote surgery systems to constellations of orbiting satellites all require enormous numbers of communicating and interacting entities. These entities must work together harmoniously to be effective. As the numbers of these interacting entities increases, ensuring their efficient operation becomes increasingly difficult. Indeed, for the past three decades this growth in general has approximately doubled every 18 months [1, p.32]. New paradigms of modern warfare also indicate an accelerated growth in the numbers of interacting systems. Amidst this growth, there is a growing consensus among experts that current network management approaches will be insufficient to handle the level of complexity that is envisioned [1] [2]. Consequently, new approaches for network management and control in complex systems are needed.

---

[1] As stated by the National Research Council in a report on countering terrorism: "Research is also needed for self-adaptive networks that can reconfigure themselves in response to damage and changes in demand, and that can degrade gracefully."[2]

One promising approach is based on what is often referred to as *Swarm Intelligence* (SI). The term SI has come to represent the idea that it is possible to control and manage complex systems of interacting entities even though the interactions between and among the entities being controlled is, in some sense, minimal. This notion therefore lends itself to forms of distributed control that may be much more efficient, scalable and effective for large, complex systems.

The underlying features of SI are based on observations of social insects. Ant colonies and beehives, for example, have the interesting property that large numbers of them seem to conduct their affairs in a very organized way with seemingly purposeful behavior that enhances their collective survival. Surprisingly and paradoxically, these insects seem to utilize very simple rules of interaction. This phenomenon is very similar to those addressed in other domains of inquiry involving *complexity* such as cellular automata and the study of chaos [3], [4], [5]. These areas along with SI have perplexed a large number of scientists for many years [6]. How is it that "swarms" of creatures with relatively low brain power and communications capabilities can engage is what is often termed "emergent behavior" reflective of some "collective intelligence" [6, p.6] —behavior that seems to exhibit a more global purpose?

Unfortunately, there is no widely agreed upon definition of what SI is or how it should or could be mathematically defined or characterized. Many terms have been associated with SI such as *emergent behavior*, *self-organized behavior*, *collective intelligence*, and the like and have been used in a variety of contexts and associated with a host of applications [7], [8], [9], but these terms also suffer from vague definitions or descriptions. There is no general, mathematically oriented description that ties all of these concepts together.

The lack of precise definitions and, hence, theoretical foundations, poses a number of significant problems and even causes confusion. The lack of precise definitions is the least of the problems—this confusion also entails missed opportunities as well. All these different descriptions and implementations muddy the waters of how to productively utilize SI concepts. Without a clear understanding of what SI is and how and why it arises, it is very difficult to envision how to take advantage of its true potential.

The remedy for this apparent confusion comes from new perspectives that illuminates the *fundamental* properties of SI. This article seeks to do just that by articulating some useful ideas based on perspectives from evolution, notions

of efficiency, and adaptability coupled with a more formal definition of self-organized behavior.

This article formulates three strategic or foundational frameworks from which to build a successful theory of SI and which provide guidelines for how to implement SI concepts. These frameworks are based on the development of the successful theories pertaining to the *simulated annealing* (SA) algorithm. The application of these frameworks to SI may provide the necessary yet still missing ingredients for developing a better understanding of SI. This may lead to entirely new ways of viewing and understanding this paradigm and ultimately allow for more practical implementation schemes.

The most important part of this foundational triad, and a necessary ingredient for developing a solid theoretical foundation for SI, is the articulation of some *first principles* based on the relevant laws of nature and their implications. This serves as a guide that governs and constrains the articulation of the other important components in developing a useful theory. For SI, these first principles are based on the laws of evolution and natural selection. As the reader will discover, its reasonable implications suggest a more precise and mathematically useful definition of self-organized behavior—*the evolution of system states along a Pareto optimal frontier*.

The second component of this triad is the articulation of an appropriate *dynamical framework*, a way of characterizing system dynamics that is a consequence of or constrained by the first principles. The dynamical framework suggested here is based on a new concept in the context of SI— *Scale Invariant Pareto Optimality* (SIPO). SIPO is a powerful concept that captures notions of symmetry and scale invariance and can address the issues of how swarms of entities communicate, modify their behavior, and *adapt* to changing environmental conditions—one of the hallmarks of SI. This dynamical framework also provides a set of rules that, in effect, imposes constraints on a system's dynamics so as to maintain consistency with the implications of the relevant laws of nature as articulated in the first principles.

Finally, the third component of this triad is the articulation of an appropriate *problem framework*. This provides useful ways to abstract these ideas and allows them to be implemented. It provides a concrete way of defining problems that help to further narrow the issues and focus research and development efforts. This article describes a general test-bed approach using the concept of *swarming finite-state machine* (SFSM) models.

Together, this triad of frameworks, or *meta-formalism*, present a unified scheme for approaching the research problems and investigating ways to implement SI concepts. It addresses how swarms of entities must communicate and modify their behavior in response to information from other entities and their environment for there to exist the emergent, self-organized behavior known as "swarm intelligence". This set of perspectives leads to a more precise mathematical definition of SI, describes ways to more fully take advantage of this paradigm and, constructively address any of its inherent limitations.

The rest of this article is organized as follows: Section II provides more detail on the current research environment regarding SI and its myriad of applications. Section III describes why the three formalisms described above provide a sound basis for developing the theoretical foundations of SI by describing similarities to the *de facto* frameworks associated with simulated annealing. Sections III-A through III-C describe the three frameworks on a conceptual level. Finally, Section IV provides concluding remarks.

II. BACKGROUND: THE SWARM INTELLIGENCE PARADIGM

A. Observations of Social Insects

Observations of social insects such as ants and ant colonies provide a great deal of insight into their behavior and SI in general. Ants and ant colonies have several ways of solving different but related problems. The main mechanism for solving them is through the use of chemical substances known as *pheromones* which have a scent that decays over time through the process of evaporation [6, p. 26]. These pheromones form the basis of what amounts to a clever, and apparently simple, communications and information storage and retrieval system. Since pheromone strength or intensity decays over time, it also provides a very simple information processing mechanism that can implement forms of positive and negative feedback [6, pp. 9-10, 41] and *reinforcement learning* mechanisms [6, p.96]. This "processing" capability is illustrated in the simplicity of how ants utilize and respond to pheromones.

As an example, consider how ants actually solve shortest path problems. Their motivation for solving these problems stems from their need to find sources of food. Efficiency dictates that they find sources closest to their colonies. Ants (many ants) first set out in search of a food source by randomly choosing (apparently randomly) several different paths. Along the way they leave traces of pheromone [6, p. 42]. Once ants find a food source, they retrace their path back to their colony (and in so doing inform other ants in the colony) by following their scent back to their point of origin. Since many ants go out from their colony in search of food, the ants that return first are presumably those that have found the food source closest to the colony or at least have found a source that is in some sense more accessible. In this way, an ant colony can identify the shortest or "best" path to the food source [6].

The cleverness and simplicity of this scheme is highlighted when this process is examined from what one could conceive of as the ants' perspective—they simply follow the path with the strongest scent (or so it seems). The shortest path will have the strongest scent because less time has elapsed between when the ants set out in search of food and when they arrive back at the colony, hence there is less time for the pheromone to evaporate. This leads more ants to go along this path further strengthening the pheromone trail and thereby reinforcing the shortest path to the food source and so exhibits a form of reinforcement learning [6], [10], [11].

But this simple method of reinforcement or positive feedback also exhibits important characteristics of efficient group behavior. If, for instance, the shortest path is somehow obstructed, then the second best shortest path will, at some later point in time, have the strongest pheromone, hence will induce ants to traverse it thereby strengthening this alternate path.

Thus, the decay in the pheromone level leads to *redundancy*, *robustness* and *adaptivity*, *i.e.,* what some describe as *emergent* behavior [6].

Many optimization algorithms attempt to imaginatively capture some notion of SI. Indeed, many difficult optimization problems have been solved by so-called *ant algorithms* such as the Traveling Salesman Problem, the Quadratic Assignment Problem and other *NP*-hard optimization problems (see [6] for a large number of examples and citations). These algorithms generally utilize some analogue of pheromone or some simple stigmergic signalling mechanism. AntNet [12] for example, uses reinforcement learning to increase the probabilities of using certain routes in a routing algorithm. The probability value is used as an analogue to pheromone. Another example is in [13] which uses a similar update mechanism to control unmanned aerial vehicles. These different approaches all try to take advantage of how social insects seem to function. These attempts to implement some SI characteristic however often are forced to creatively sidestep the concept of self-organization and its implications.

B. The Mystery of Self-Organization

Although much has been learned from observations of these social insects, there is no widely agreed upon definition of what constitutes SI. Indeed, the term *SI* is bandied about so often and in such a wide variety of contexts that it causes confusion leading to many different interpretations and implementations for various problems. Although many of these implementations reflect some notion of SI, they often entail other paradigms as indicated above such as *reinforcement learning* [11], [14] and *stigmergy* which refers to complex *indirect* interactions based on simple signalling systems [6, p.14]. Where does one paradigm end and the other begin? Still, other research seems to freely use the term SI when there simply are large numbers of interacting entities (see [15] for a complete discussion on competing paradigms). Does that alone suffice to describe or define SI? Is it merely a way of somehow taking advantage of parallel computing methods?

Notwithstanding the many different descriptions offered by many researchers, the main features of SI seem to involve forms of limited or minimal communications and/or interactions, large numbers of interacting entities with limited reach, and some globally efficient, emergent or *self-organized* behavior [6, p.9]. Bonabeau *et al.* [6, p.9-11] suggests that the central features of SI are based on the manifestation of *self-organization* that arise from the interplay of four basic ingredients: 1) forms of positive feedback, 2) forms of negative feedback, 3) the *amplification of fluctuations* that give rise to structures, and finally 4) multiple interactions of multiple entities.

But even this characterization provides very little insight into what SI is except in very descriptive terms. For example, it does not fully explain in a unifying way why or how pheromones evolved in the way they did, or why they should have the evaporative properties they have, or why they have their chemical makeup, or how ants can somehow distinguish among different types of pheromone (presumably in order to

identify ants from other colonies) [16, p.156]. Pheromones are complex chemical signalling systems, yet most of the research that deals with them or models their effects use the concept in very limited ways, *e.g.,* as a scalar in ant algorithms [14], [17] as opposed to a more complex scheme represented by vectors. Although, as we shall see, even simple scalars can possess enough information related to a measure of efficiency, pheromones are likely to have more complicated properties than their mere intensity [6]. Indeed, Wiener [16] in his ground-breaking book *Cybernetics* emphasizes the importance of intercommunication among the entities in question:

> How then does the beehive act in unison, and at that in a very variable, adapted, organized unison? Obviously, the secret is in the intercommunication of its members. . . This intercommunication can vary greatly in complexity and content . . . the value of a simple stimulus, such as an odor, for conveying information depends not only on the information conveyed by the stimulus itself but the whole nervous constitution of the sender and the receiver of stimulus as well [16, p.156-7].

However SI is described, one of its central characterizations is that of *self-organization*. But this also begs the question of what constitutes SI because there is no clear understanding of what self-organization or emergent behavior is! These terms have been around for some time and their definitions have been and probably will continue to be debated for some time.

Serra *et al.* describes the concept of self-organization generally as "highly organized behaviour even in the absence of a pre-ordained design." [18, p.1] (but what is 'organized'?). They go on to further describe examples such as the resonance phenomenon in lasers, and in cellular automata where "unexpected and complex behaviours [] can be considered as self-organized." [18, p.2].

Earlier, Prigogine[2] *et al.* [5] described self-organization in chemical reactions and thermodynamic contexts. His notion is quite enlightening and emphasizes the element of *fluctuations* in far-from-equilibrium system states. Non-linear system changes, fluctuations as he refers to them, due to either random events or chaotic dynamics, are then amplified by positive feedback mechanisms. This results in structures that emerge spontaneously which often are presented as examples of self-organization. It is interesting that Prigogine provides an example of SI based on the clustering behavior of termites in constructing termite nests [5, p.181-186]. See also [6, p.207-234]. Even earlier, Wiener [16] used the term in describing the brain waves of humans. See [16, Chap. 10: Brain Waves and Self-Organizing Systems]. Again, the notion of what self-organization is remains unclear. Is it merely some structure or pattern? And if so, how is one to distinguish it from merely random effects?

What is needed to truly take advantage of SI is more than the mere descriptions of the attributes of SI and self-organization. Something amenable to formal mathematical definition would be quite valuable. This article offers a new and potentially

---

[2]Prigogine won the Nobel Prize in 1977 for his work on the thermodynamics of non-equilibrium system.

useful working definition of emergent or self-organized behavior for SI based on the imperatives of evolution and natural selection (as opposed to the imperatives of thermodynamics and irreversibility that elicited Prigogine's definition [5]). The concept of self-organization therefore plays a central role in the development of the foundational principles of SI.

These principles are founded on the notion that self-organized behavior of the type observed in entities subject to the laws of evolution involves forms of efficiency in resource allocation. It emphasizes the importance of *signals* as mechanisms of change in system states subject to the imperatives of the evolutionary pressures of natural selection. It is important to recognize that the response of insect swarms to external stimuli is governed by processing systems that have been heavily influenced and affected by the forces of evolution. This allows consideration of a much richer spectrum of behaviors, both simple and complex, than those implied by the mere application of positive and negative feedback mechanisms to simple signals. In fact, the type of self-organization described here can be framed as *a form of symmetry* in that changes in the system's behavior and operating points nonetheless leave unchanged certain attributes associated with these operating points, namely, their Pareto optimality [19]. In short, a novel definition of self-organization presented here is *system behavior that maintains its operating points on or near a Pareto optimal frontier.* This notion of efficiency constitutes a central feature of the foundational principles for SI and is described in greater detail in the next section.

III. THE IMPORTANCE OF BASIC PRINCIPLES

Developing a viable set of basic principles underpinning SI concepts is essential for successful implementations. The principles described here have a certain credibility and hence viability because they are based on how other successful theories appear to have been developed. This requires some discernment and creative articulation of the main features of how these other theories were, in fact, developed and how these features could be applied in the case of SI. The simulated annealing (SA) algorithm is a good example of this process and provides clues on how to articulate the corresponding components of the central principles underlying SI.

A great deal of theoretical results on SA have been published in many papers and books (see *e.g.,* [20], [21], [22], [23]). But this algorithm was not developed or analyzed in a vacuum. Its development was, in fact, based on the confluence of certain factors. Three main factors or components of the research evolution are discernable. The first component consists of a set of *first principles* based on the *relevant laws of nature and their implications*. The second component is an appropriate *dynamical framework*. Finally, the third component is an appropriate *problem framework*. These three components form the core of what can be referred to as a set of foundational frameworks. To see how these components provide guidance in developing a viable research strategy for SI, an understanding of how they worked together in the case of SA is helpful.

In SA, the first principles were based on the laws of thermodynamics. The most salient implication of this law was that the entropy of a thermodynamic system, the randomness of its states, tends to its maximum value at any given temperature. Metropolis [24] utilized this idea in developing a method for simulating the evolution of a complex thermodynamic system. The underlying law of thermodynamics was manifest in the objective function of a non-linear program (NLP). The objective of this NLP was to maximize the entropy of a stationary Markov chain subject to certain constraints (such as the the constraint that expected energy level is proportional to temperature) [21].

The solution of this NLP lead to an expression for the stationary probability distribution (the Boltzmann Distribution) of each system state [21]. Thus, the implication of the first principles was the expression of the Boltzmann Distribution. What remained was to somehow develop an appropriate mechanism that described how this thermodynamic system changes from one state to another while at the same time remaining consistent with the maximization of entropy, *i.e.,* the Boltzmann Distribution, as dictated by the fundamental laws of thermodynamics.

Thus, in addition to this fundamental law, a *dynamical framework* was necessary to address the problem of how to model a system that undergoes some type of change. The well-known mathematical framework of Markov chains and its global and detailed balance equations provided those necessary elements that mathematically describe the stochastic nature of the system as it changes from one state to another while still obeying the fundamental laws of thermodynamics [20], [22], [24]. It was this utilization of the detailed and global balance equations that ultimately lead Metropolis *et al.* to his now famous *Metropolis Acceptance Criterion*, the main result of Metropolis' contribution. In effect, Metropolis 'engineered' a transition probability to fit within the framework of global and detailed balance equations given the stationary probability, *i.e.,* the Boltzmann Distribution.

Once these transition probabilities were defined, the mathematical description of how a large, thermodynamic system evolves from one state to another in a computer simulation was complete [24]. Once this method of simulating a thermodynamic system was in place (in 1953), the idea of reducing temperature slowly to "anneal", conceived by Kirkpatrick [25] some 30 years later, was the final element that lead to the SA algorithm. This algorithm has since been fully mathematically described and analyzed by a host of researchers[3] (see [20], [22], [24], [25]) using various *combinatorial optimization problems* (COPs) as the *problem framework*—a general way of characterizing a whole host of discrete optimization problems. Thus, the dynamical and problem frameworks provided the mathematical constraints and research focus in a way that was consistent with the laws of thermodynamics.

To summarize how these three components of the foundational principles, the meta-formalism, describes the evolution

---

[3]Although the algorithm seems to have been fully analyzed, this proposer believes there are still a number of fundamental theoretical elements that await discovery. See *e.g.,* Fleischer [23] for a description of scale invariant properties of the SA algorithm. These properties illustrate some potentially important symmetries in SA and may play a role in research involving a group theoretic approach for modelling SI.

of SA theory, the first principles were based on thermodynamics manifest as the maximization of the entropy in an NLP that lead to the definition of the Boltzmann Distribution. The dynamical framework was based on the global and detailed balance equations of Markov chains which lead to the Metropolis Acceptance Criterion and the transition probabilities in SA. Finally, the problem framework was based on COPs which permitted the application of the ideas onto test bed problems for experimentation and analysis. A successful and useful set of formalisms can therefore be modelled on this developmental process[4] hence ought to involve first principles based on fundamental laws, an appropriate dynamical framework, and an appropriate problem framework. Each of these three components can be recast to within the context of SI. This is described in more detail in the next three sections.

A. First Principles: The Laws of Evolution

Any successful theory requires some reasonable and widely accepted assumptions on which the theory is based. Because the SI paradigm is based on observations of social insects, *i.e.*, living creatures, the first principles ought to based on, involve, or at least be consistent with, the laws of evolution and its implications. Basically, this theory states that species evolve over generations constantly improving their fitness to survive through the process of natural selection and genetic mutation. Indeed, this theory forms the basis of GAs and its associated theoretical results [27], [26]. But this theoretical statement by itself seems insufficient to establish foundational principles for SI. Some *reasonable implications* of this theory are needed.

One reasonable and important implication of the theory of evolution as it relates to swarms of insects is that it selects out those systems that exhibit *efficient behavior on many scales*. This notion of efficiency is not only reasonable, but is also based on observations of social insects and how they *adapt* to changing environmental conditions (see [26]). Indeed, it seems basic that an efficient allocation and use of resources provides a distinct survival value to any species. Thus, a reasonable implication of evolution regarding the behavior of social insects is that it lead to behaviors that are, in some measure, *efficient*. These notions of efficiency and adaptability have an important mathematical formalism based on the concept of *Pareto optimality*.

1) Efficiency Via Pareto Optimality: Optimization problems are ubiquitous in the real world and social insects must also deal with a variety of them if they are to survive. Certainly, the efficient allocation of resources present problems where some goal or objective must be maintained or achieved. Such goals or objectives are often mathematically modelled using *objective functions*, functions of decision variables or parameters that produce a scalar value that must be either minimized or maximized. The challenge presented in these often difficult problems is to find the values of those parameters that either minimize or maximize, *i.e.*, optimize, the objective function value subject to some constraints on the decision variables.

Most complex systems, however, have *several* objectives that must be considered when assessing overall system performance. Operations of these systems sometimes leads to conflicting objectives where one objective must be *efficiently* "traded off" for another. In these *multi-objective optimization problems* (MOPs) system efficiency in a mathematical sense is often based on the definition of *Pareto optimality*–a well-established way of characterizing a *set* of optimal solutions when several objective functions are involved (see *e.g.,* [28] and the citations therein).

In this perspective, each operating point or vector of decision variables (operational parameters) produces several objective function values corresponding to a single point in objective function space (this implies a vector of objective function values). A *Pareto optimum* corresponds to a point in objective function space with the property that when it is compared to *any* other feasible point in objective function space, at least one objective function value (vector component) is superior to the corresponding objective function value (vector component) of this other point. Pareto optima therefore constitute a special subset of points in objective function space that lie along what is referred to as the *Pareto optimal frontier*–the set of points that together dominate (are superior to) all other points in objective function space [29]. See also [30] and [31] for descriptions of multi-objective GAs (their intriguing titles notwithstanding do not encompass the ideas put forth here on SI).
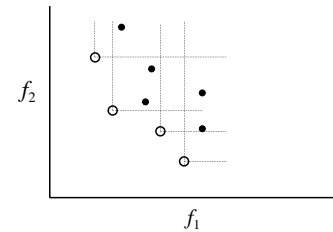

Fig. 1. The Pareto Optimal Frontier

An important characteristic of Pareto optima, this frontier of points, is that they correspond to operating points where the improvement of one objective function value comes only at the cost of worsening some other objective function value— trading off one for another. Figure 1 illustrates this set of points in objective function space by the open O's where each operational decision vector produces two objective function values f_1 and f_2 that are minimized. It is along this set of points that operational decisions must be restricted if operational efficiency is to be maintained. It is easy to see that the Pareto frontier dominates (is superior to) the other points in this 2-dimensional objective function space.

Effective methods for determining several Pareto optima can be quite valuable for enhancing the survival value of a species (or managing a complex system) because it enables *adaptive* behavior. This allows the possibility of efficient operations when there are changes in the relative utility of the several objectives or when one objective must be restricted to a certain range of values. Under such circumstances, the system can be

---

[4]We believe that other theories such as those associated with genetic algorithms (GAs) offer a similar developmental model.

moved along the Pareto frontier from one Pareto optimum to another thereby maintaining operational efficiency. Thus, if in an ant colony a path to a food source becomes congested, then other routes must be utilized. Although the distances to food sources are generally minimized as is the level of congestion, these often conflicting objectives can be efficiently traded off when the shortest distance is sacrificed to lessen the level of congestion [6].

To summarize this first component of the foundational principles of SI, the first principles and its implications suggest that both subsystems and system should behave in a manner consistent with Pareto optimality. What remains is to characterize *Pareto optimal behavior* in a context reflective of SI. It is noteworthy, and somewhat surprising, that these ideas pertaining to Pareto optimality and self-organization have apparently not been pursued or even articulated by others in the context of SI research! See *e.g.,* the very complete books on SI such as [6]. Assuming that natural selection pressures have imposed behaviors leading to efficient, adaptable, *i.e.,* Pareto optimal behavior, the question of how this is achieved in social insect societies must be addressed. What are the mechanisms that social insects use to solve what amount to MOPs with the limited forms of communication they seem to have?

It may be that notions of scale are involved. Indeed, the concept of *stigmergy* suggests that what has been regarded as collective intelligence or emergent behavior on a large scale is effected by decisions on a smaller scale [6, p.14]. But how is such a relationship among different scales possible when decisions affecting the local environment are based on apparently *simple* information? How can system change be mathematically described in a way consistent with movement along the Pareto optimal frontier? More significant however is the notion of *efficiently trading off the efficiencies on one scale with efficiencies on another scale*. Addressing these issues mathematically is the purpose of the dynamical framework described below.

B. The Dynamical Framework: Promoting Adaptability

The purpose of the dynamical framework is to provide guidelines for how to mathematically model the dynamics of the system in a way that is consistent with the first principles and faithfully reflects the desired phenomenon. Its articulation provides the necessary research and development focus for describing system dynamics. This allows *specific questions to be asked* that lead to specific tests, mathematical constraints, and the like. In essence, it facilitates the *engineering* of the dynamical rules that, in this case, lead to Pareto optimal behavior. It is the essential feature that allows this type of research to proceed.

For SI, the dynamical framework must be based on the two intertwined principle features of SI—its emergent or self-organized behavior, and the minimal information processing and communications capabilities of the entities involved. These two rather vague notions must somehow lead to more well-defined properties and characterizations. The preceding section suggests that the emergent behavior can be characterized by Pareto optimal operating points in some decision

space. The additional requirement that such efficient behavior arises from limited communication and information, in effect, imposes some scaling aspects. These are addressed by the *Scale Invariant Pareto Optimality* (SIPO) concept described in the next section.

1) Scale Invariant Pareto Optimality: The SIPO concept stems from a belief that evolution has lead to behaviors of social insects that promote the entire species. Somehow over the eons of evolution, insects societies have developed rules of conduct that are biologically driven and lead to efficient management of their environment and resources [6, p.30]. Yet, given their apparently limited capacity for communication and information processing, it seems rather mysterious that in spite of these limitations, their behaviors seem almost guided by an invisible hand[5] that somehow efficiently manages an entire colony's resources! How is this possible? As stated earlier, system movement along the Pareto optimal frontier lends itself to adaptable and efficient behavior thereby increasing the survival value of these social insects. The heart of the matter in SI therefore becomes:

> How can an entire system be moved along the Pareto optimal frontier when the entities that comprise the system have only limited information available to them on which to base their own behaviors and where their behaviors have only limited effects on their environment and other entities?

The SIPO concept attempts to answer these types of questions. Simply put, it is a way of characterizing Pareto optimal solutions on many scales and thus reflects efficient and adaptable system behaviors on many scales. It provides a basis for defining properties of systems that are reasonable and have been observed in social insect colonies. Evolution can be seen as imposing not just behaviors that enhance survival of individual entities, but also of groups or societies of these entities [6, p.30-1]. Thus, while ants must manage their own survival, they also derive distinct advantages from participating in a colony. Indeed, scientists have observed behaviors in ant colonies that in many respects is efficient on several different scales involving sub-societies or sub-colonies (see [6, p. 209] for a discussion on nest modularity). There are divisions of labor, aspects of specialization and such that contribute to the overall survival value of an entire colony [6, p.2]. This interdependency seems to promote their collective, and hence individual, survival.

Since the SIPO concept revolves around the concept of Pareto optima[6] it is necessary to fully characterize its mathematical attributes. So far, only its definition has been described. The next few paragraphs describe another important aspect of Pareto optima.

---

[5]Their economies like those of humans may indeed be guided by Adam's Smith's *invisible hand, i.e.,* swarm intelligence!

[6]It is worth noting that this description of SIPO concepts seems, on the surface, to be similar to those articulated by others. Menczer [31] *e.g.,* describes an evolutionary algorithm that seeks to determine Pareto optima and has some scaling properties. But this article does not describe the adaptable behavior of entities in terms of movement along a Pareto frontier. Coello *et al.* [30] also provides an intriguing title, but their paper also deals with evolutionary optimization schemes for solving MOPs and does not address issues pertaining to SI.

a) The Measure of Pareto Optima: A rather intuitive yet surprisingly little known aspect of Pareto optima is its measure. This measure is based on the size of the set of points in objective function space that are dominated by the Pareto optimal frontier—in essence a Lebesgue measure or hypervolume. Zitzler [32] was apparently the first to publish this idea in a brief three sentence description.[7] Fleischer [28] extended this idea and described a set function that maps a set of points in objective function space to a scalar value, the Lebesgue measure, and formally proved that this scalar achieves its maximum value if and only if its arguments are Pareto optima. Fleischer also defined an efficient (polynomial) algorithm for calculating this scalar for an arbitrary number of dimensions (objective functions).
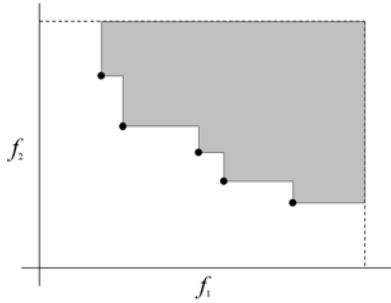

Fig. 2.   The hypervolume of Pareto optima.

This hypervolume is illustrated by the shaded region in Figure 2 for two minimizing objective functions f_1 and f_2 (see [29] for more complicated illustrations in 3 dimensional space). The dotted lines are upper bounds on f_1 and f_2 while the black dots correspond to feasible points in the underlying decision space. Because this set function produces a scalar, it can be used as an objective function in an optimization algorithm such as SA. The proof that if this hypervolume is maximized then its arguments are Pareto optima[8] implies that its use in SA can induce it to converge in probability to the Pareto optima![9]

One important aspect of this measure is that it *quantifies efficiency, i.e.,* changes in the overall efficiency can be mathematically captured by a changes in the scalar hypervolume and depicted by a receding or expanding Pareto optimal frontier (see below for a further discussion on efficiency tradeoffs). This measure and its potential roles in SI are described in the next sections.

b) Subsystem Interactions and Constraints on Objective Functions: The SIPO property suggests that local efficiency

may often be sufficient for *system-wide* efficiency. Articulating and applying this concept to systems of entities therefore imposes constraints on how subsystems can interact, *using simple forms of information,* so that a system, be it an ant colony, a swarm of chemical sensors, or a network of computer systems, operates on, or close to, the Pareto optimal frontier on both local and global scales. How this can be accomplished is non-trivial because several identical subsystems may operate on *different* Pareto optima in terms of local performance metrics.

Consider the following simple example where two, identical subsystems have operating points (vectors of parameters or decision variables) on one of two local Pareto optima $\mathbf{x}_1$ and $\mathbf{x}_2$. These two operating points lead to points in objective function space $\mathbf{p}_1 = (f(\mathbf{x}_1), g(\mathbf{x}_1)) \equiv (f_1, g_1)$ and $\mathbf{p}_2 = (f(\mathbf{x}_2), g(\mathbf{x}_2)) \equiv (f_2, g_2)$ for two local objective functions f and g, where, for purposes here, both are minimized. Let functions F and G be global objective functions each taking two arguments, one from each subsystem. Assuming the order of arguments makes no difference in these functions, there are three distinct ways in which these two subsystems can operate on local Pareto optima: two ways where both subsystems operate on the same point in local objective function space, *i.e.,* $(F(\mathbf{x}_1, \mathbf{x}_1), G(\mathbf{x}_1, \mathbf{x}_1))$ and $(F(\mathbf{x}_2, \mathbf{x}_2), G(\mathbf{x}_2, \mathbf{x}_2))$ and one way in which the subsystems operate on different local Pareto optima, *i.e.,* $(F(\mathbf{x}_1, \mathbf{x}_2), G(\mathbf{x}_1, \mathbf{x}_2))$. How should the global objective functions F and G be defined in terms of local objective functions so that all three combinations of local Pareto optima map into the set of global Pareto optima?

More generally, under what circumstances do different *combinations* of local Pareto optima associated with subsystems yield a global Pareto optimum? Put another way, what is the form of objective functions that allow mappings of Pareto optimal points on the sub-system level, to Pareto optimal points on the system-wide level? This scaling from the sub-system level to a system-wide level may therefore impose certain additivity properties or exponential forms in the measures of performance and presents separability issues. For example, when must $F(\mathbf{x}_1, \mathbf{x}_2) = f(\mathbf{x}_1) + f(\mathbf{x}_2)$? This is another example of how the SIPO concept narrows down the playing field and sharpens the questions that need to be asked and answered in any development work involving SI concepts.

Although it is desired that subsystem states be maintained along a local Pareto optimal frontier at the same time system-wide states are maintained on a global Pareto optimal frontier, the possibility exists that this may not always be possible depending upon how these objective functions and their relationships are defined. Afterall, subsystem states affect the system-wide objective functions which are also affected by large numbers of other subsystems. Moreover, subsystems do not behave independently. The resulting complexity is similar in many ways to the complexity of cellular automata (see below). The following section describes how can these constraints can be captured in a mathematically useful way that still reflects efficiency.

c) Efficient Tradeoffs of Efficiency Between Scales: Sometimes the benefits to the individual entity must be traded

---

[7]This idea was independently formulated by the author who later discovered that Zitzler had already described it in a brief three sentence passage. The author was therefore compelled to provide a formal proof and some algorithmic extensions.

[8]This direction of proof of the *if and only if* implications is much more difficult, but is necessary in order to justify its potential use as an objective function in a metaheuristic such as SA where the mechanism of search is crucially dependent on the objective function value, *e.g.,* the Metropolis Acceptance Criterion the goal of which is to improve the objective function value.

[9]The author is not aware of any other multi-objective optimization scheme that theoretically converges in probability to Pareto optima.

off for the good of the entire group—*e.g.,* soldiers must often make the ultimate sacrifice to ensure the success of their unit (or nation).[10] This suggests the possibility that certain states of neighboring subsystems maybe 'incompatible'. That is, a certain Pareto optimum in one subsystem may preclude a particular Pareto optimum in another (similar to Ising spin glass models—see the section below on Markov Random Fields). Thus, it is conceivable that only certain combinations of Pareto optima among neighboring subsystems may coexist at any given time. If this is true, it may explain the notion of *fluctuations* as Prigogine calls them [5] and the formation of structures as an attribute of self-organization.

If such incompatibilities exists between and among neighboring subsystems or between subsystems and system-wide Pareto optima, then it must be the case that there is some trade-off between the efficiency measures of individual subsystems and/or the efficiency measure of the entire conglomeration of entities. Some subsystems may have to sacrifice efficiency in order for other subsystems to operate Pareto optimally. Recall from the earlier discussion that the measure of Pareto optima is a quantification of the efficiency of a system. Thus, a mathematical way of characterizing a tradeoff of efficiency between two subsystems or between two scales of systems is to use their respective measures of efficiency to define a Pareto optimal frontier based on them. In effect, **the scalar hypervolumes associated with Pareto optima on different scales themselves become objective functions for a Pareto optimal frontier!** To the author's knowledge, this type of tradeoff has never been characterized or described in this fashion, hence provides a novel way of characterizing subsystem to system interactions. Figure 3 depicts such a tradeoff curve where the Pareto optimal frontier is the bolded part of the curve. Note that improvement on one scale comes at the cost of worsening the hypervolume on the other scale. Note also the concave form of the curve which is based on the ideas that follow.[11]

d) Structures and Self-Organized Behavior: The descriptions of self-organized behavior described earlier often entail the notions of structures or patterns as the main characterization of self-organization. The foregoing discussion on efficient tradeoffs of efficiency measures provides another way of describing how such structures and their precipitating fluctuations arise.

First, some random configuration among the subsystem or entities occurs. Next, the subsystems all seek their local Pareto optimum. Finally, in the case where some subsystems cannot attain a Pareto optimum in terms of local objectives because of state incompatibilities among neighboring subsystems, these subsystems attempt an efficient tradeoff of efficiency measures as described above. These tradeoffs are between an estimate of the measure associated with the global Pareto frontier and an estimate of the measure associated with the local

Pareto frontier. Thus, as the subsystems and system evolves, neighboring subsystems may evolve in a way guided by these tradeoffs of efficiency measures. Thus, the state incompatibilities ultimately give rise to structures as subsystems attempt to operate on their local Pareto optimal frontier. The patterns of states are therefore governed by their initial conditions, the various state incompatibilities in terms of Pareto optima, and the efficient tradeoffs of efficiency measures in the midst of these subsystem state incompatibilities. The patterns or structures observed in insect swarms can therefore possibly be explained in these terms.

e) The 'Graceful' Degradation of Performance: The foregoing discussion also suggests a general and mathematical way of describing the "graceful degradation" of system performance. Such graceful degradation has been seen as an important component in the management of large scale systems such as the Internet [2, Ch.5]. Rather than suffering catastrophic changes to a system, it may be possible for system efficiency to be degraded in a more gradual manner. This is because system efficiency can be measured using the Lebesgue measure described earlier, hence the measures for different scales can be traded off efficiently! This idea of efficient tradeoffs of efficiency measures provides a *quantifiable guide* of how best to achieve this sacrifice of efficiency on one scale for improved efficiency on another scale. These types of issues can arise in a variety of contexts.
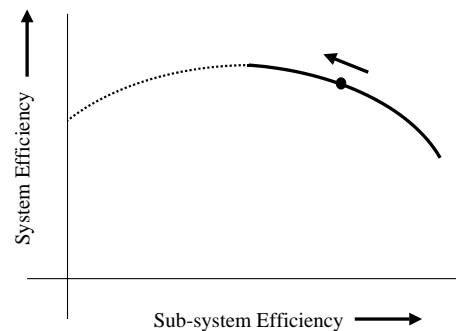

Fig. 3.   A tradeoff curve of different measures of efficiency.

For instance, during the operation of a complex system it may happen that environmental changes, or changes in system state results in a subsystem-system configuration depicted in Figure 3. Here, the $x-$axis is an estimate of a particular subsystem's efficiency measure and the $y-$axis is an estimate of the global efficiency measure. The black dot corresponds to the subsystem/system efficiency point under current environmental conditions. These conditions suggest a tradeoff of efficiency is possible where the system's efficiency can be improved at a cost of decreasing the subsystem efficiency, *i.e.,* the black dot is moved in the direction of the arrow.

Obviously, if the efficiency measures of enough subsystems are degraded, it becomes increasingly difficult to compensate for their degradation. Eventually, the entire system's efficiency measure will be reduced, but the actions of the subsystems in establishing their movements along this Pareto optimal frontier of these efficiency measures, provides for the most graceful

---

[10]This type of behavior has also been observed in insect colonies where soldier ants and worker bees sometimes sacrifice their lives for the good of the entire colony. See [6, p.36-9].

[11]It should be borne in mind that the two axes depicted, one for subsystems and the other for the entire system are not independent, but this simple curve captures the essential idea of the efficient tradeoffs of efficiency measures.

degradation of efficiency—because the degradation occurs efficiently! How this occurs poses an interesting and, needless to say, difficult problem owing to the many subsystem interactions with the system, *i.e.,* the complexity problem and the fact that these efficiency measures are not independent. Without be able to quantify efficiency with this scalar hypervolume, however, it would be impossible to address this problem of graceful degradation in any mathematical way that provides some guidance and direction for how to achieve this. This approach of characterizing the subsystem to system efficiency tradeoff therefore has many applications to both information networking, understanding the economics and management of such systems, as well as economics in general!

SIPO is a new way of articulating emergent behavior in the context of systems operating under the imperatives of natural selection. It provides a way to mathematically characterize properties of systems on many scales and imposes constraints on how the subsystems interact and functionally affect global measures of performance. This makes implementation of efficient systems more realistic, provides some underlying structure from which to investigate mathematical forms for objective functions, has the appealing feature of scale invariance and makes sense from an evolutionary point of view. But can such properties arise in systems with limited forms of communication and information processing? The following section provides a clue, based on recent research, that says it is indeed possible for Pareto optima to at least be associated with limitations on the form of information.

2) Limited Signalling Systems and Pareto Optimality: The nexus of operational efficiency and SI seems to stem from how the swarms of entities, *i.e.,* sub-systems, communicate and interact with each other and their environment. If we are to assume that biological systems have evolved over time to operate efficiently, then the signalling systems used by social insects must have some mathematical connection to multi-objective optimization and the concept of Pareto optimality. Moreover, such a signalling system must be simple in some sense. The SI paradigm as indicated in the literature is based on the use of very simple (and apparently low-cost) signalling methods. Study of ant colonies for example, shows that entire social systems operate efficiently using simple chemical signalling systems based on pheromones. These pheromones constitute a conceptually simple signalling system and provide a great deal of useful information, *e.g.,* temporal information since the strength of the pheromone decays over time by the process of evaporation ([6, p.43]) in addition to memory of where the ants have been. By interpreting these chemical signals and responding to them with simple behaviors, ants perform complex functions efficiently.

The idea of using a simple signal to convey information that leads to efficient behavior is somewhat remarkable and the question therefore arises as to whether a simple signal can even be associated with Pareto optima. The answer to this question is of course that it is possible in the sense that *an entire set of Pareto optima can be measured using a single scalar value*, the hypervolume or Lebesgue measure as described earlier (in the paragraph on the measure of Pareto optima). Certainly, a single scalar quantity qualifies as perhaps

the simplest signal [33]. Thus, a strong connection between Pareto optima and simple signals is possible. In effect, this scalar may constitute a mathematical analogue of pheromone! This scalar quantity not only provides the foundation for a whole host of heuristic methods that can reveal Pareto optima (or near Pareto optima, see [28], [29][12]), but may also enable the operational movement along a Pareto optimal frontier.

Pheromone may therefore be nature's way of mapping an entire set of Pareto optimal points to a single scalar quantity and suggests mathematical constraints on the form of functions that mimic pheromone. Thus, just as maximizing entropy served as a guide in developing Metropolis equations of state [24] that ultimately lead to SA, maximizing this Lebesgue measure may serve as a guide in developing appropriate signalling systems to mimic pheromone *and how other related quantities should change over time or in relation to other quantities*. This is the heart and purpose of the dynamical framework.

a) Swarm-based Solution Methods: One important consideration in this discussion on Pareto optima, efficiency and limited communications is how to tie it all together. How can one effectuate solution methods involving simple signals that, in effect, solve MOPs?

One way alluded to earlier is to use the SA algorithm. One of its parallel forms, referred to as *cybernetic optimization by simulated annealing* (COSA), uses scalars that are communicated among processors to modulate the temperature control parameter [34], [35]. These scalars provide a feedback control system that improves SAs performance. Like many ants signalling each other with simple signals, parallel processors running COSA communicate using simple signals in a way that enhances their collective performance.

This form of parallel computation as many similarities to other parallel forms such as Tabu Search and so-called "random restart local search" methods [27]. These methods all share a number of features in common with how swarms of insects solve problems *e.g.,* by exploring many different paths simultaneously (see [6, p.55]). In particular, each processor in a parallel system starts at some randomly selected point in decision space and attempts to find the global optimum from that point. With many such processors, the chances of finding the global optimum in a reasonable amount of time increases.

b) Transformations and Movement Along the Pareto Optimal Frontier: An important aspect of multi-objective optimization problems is that the vectors of decision variables that produce the Pareto optima can be scattered throughout the decision space (the domain) in an apparently random or haphazard manner [28], [29]. This is due to the interplay among the several objective functions—Pareto optima may lie on the sides of hills rather than at their tops or bottoms (the local minima or maxima). Thus, it is very difficult to determine where these decision variable are in the decision space. Some appropriate method to determine the Pareto optima and their corresponding vectors of decision variables is necessary (see [28]).

---

[12]Note that these results have also detailed *efficient algorithms* for calculating the scalar associated with this set function for any number of objective functions.

It bears emphasis that two Pareto optima that are close to one another in objective function space may have their underlying vectors of decision variables quite widely separated and vice versa. Thus, assuming that a gradual change in objective function space is desired, such as when there is a gradual change in the relative utility of the different objective functions, some method of moving from one vector of decision variables to another such vector some distance away must be devised. In other words, once the Pareto optima are identified, some method of transforming one Pareto optimal decision vector to another is required.

One approach for performing such a transformation is through the use of feedforward neural networks [36, Ch.2]. Such networks can be trained so that a Pareto optimal decision vector in the input produces the desired Pareto optimal decision vector at the output. Of course, a great deal of further research and development in this context is necessary.

This dynamical framework based on the SIPO concept and its relationships to emergent behavior, efficiency, and simple signalling systems (as suggested by the Lebesgue scalar), provide a framework that may enable novel approaches to SI research and development in ways consistent with the first principles described earlier. Together, the two main components of the SI paradigm, self-organization as indicated by movement along a Pareto optimal frontier and the minimal signalling systems as indicated by the Lebesgue measure for Pareto optima, provide a mathematical foundation from which fundamental theories and principles of SI can be developed because they are more precisely defined and characterized than the rather *ad hoc* approaches currently in vogue that attempt to capture *some* aspect of SI.

To fully accomplish this potential however, requires some way of applying these ideas to concrete and abstract problems, *i.e.,* test bed problems that facilitate the testing, experimentation and ultimately the application of these ideas. The problems used must have some potential for exhibiting emergent phenomena and at the same time allow for modelling the simple forms of interaction, the stigmergy, that is the hallmark of SI. The *problem framework*, the last of the three formalism components, provides these important elements for SI research and development and is described in the next section.

C. The Problem Framework

An important aspect of the problem framework is that it allows specific questions to be asked and problems explored. It must be sufficiently general to not only capture the essential elements of the phenomenon, but also offer some avenues on which to implement the dynamical framework. It must also be specific enough to allow useful and appropriate mathematical tools to be brought to bear. The problem framework described here attempts to achieve this by describing an abstract problem type based on what is termed here as *swarming finite state machine* (SFSMs) models.

1) Swarming Finite State Machine Models: The idea behind SFSMs is that notwithstanding the limited capabilities of swarms of entities, we should nonetheless provide some way of flexibly describing their level of complexity. Afterall,

human societies present self-organized behavior and certainly forms of 'swarm intelligence'. The allusion to Adam's Smith's "unseen hand" ([37]) and Alvin Toffler's social code based on society's propensity for concentration, centralization, specialization, standardization, synchronization, and maximization (see [38]) suggest that the central features of SI may be ubiquitous. Are these not attributes of very complex forms of swarm intelligence? What is needed then is a way to address a continuum of complexity and swarm intelligence.

SFSMs allow us to do this in that each finite-state machine (FSM) has a finite set of inputs, outputs and internal states that provide memory, the potential capacity for learning (where the transitions to states are affected by learning methods–see [39]) and certainly the complexity that we hope to manipulate and gain insight about. Moreover, the level of complexity can be changed by simply changing the numbers of states that comprise the SFSMs.

SFSMs form a type of cellular automata that when coupled with appropriate transformation functions and objective functions to reflect operation along a Pareto optimal frontier will exhibit the type of self-organized behavior defined in Section II-B. The states of each FSM will have some functional dependence on neighboring FSMs and will be designed to stabilize on efficient operating points. This may entail using genetic algorithms (GAs) or involve neural networks to search the space of transition functions or train on prescribed operating points so as to induce the FSMs to operate on Pareto optimal frontiers.

These types of simulations provide a real microcosm of the SI world. It can effectively test whether natural selection pressures are consonant with efficient operating points in the manner described here. Fitness functions related to the efficiency measures could be used to induce the SFSMs to achieve this. It would indeed be interesting to see how these SFSMs can then be designed (or evolved) to efficiently handle tradeoffs of efficiency measures, a cornerstone of the SIPO concepts described earlier.

The use of SFSMs allow for abstractions and, ultimately, practical implementations. The level of complexity can be controlled in the sense that the size of the FSMs, its number of states, inputs and outputs, can be decided upon *a priori*. The state space of the SFSMs, how its transformation functions are established and how they interact of course require further research, but this problem framework and the other two formalisms mentioned earlier provide an important structure on which real swarm intelligence can be studied and implemented for real-world problems.

It is worth pointing out, however, that much of what has been described, the properties and characteristics and such, have features in common with many other areas of inquiry in the sciences and engineering. The following areas describe some of these features that offer some potentially useful tools and ideas for developing SI systems.

D. Other Scientific/Engineering Tools

1) Group Theory: The structure of the emergent or self-organized behavior in SI based on the meta-formalism, *i.e.,*

movement along the Pareto optimal frontier, provides a number of methodologic and analytical advantages for exploring the fundamentals of SI. The most significant aspect of this structure is that it may permit the use of *group theory* in developing the fundamental theories of SI. Group theory provides powerful mathematical tools as well as the distinct possibility of developing novel insights.

This perspective requires a number of reasonable assumptions and conceptions. A basic conception is that we can model each entity i in a swarm as a *subsystem* that exists in a number of different *states* $\mathbf{x}_i = (x_{i1}, \ldots, x_{in})$ defined by n parameters corresponding to values associated with the local environment as well as the status of the agents themselves. The social insects, the ants for example, can be viewed as the *agents of change* in that they modify the states of these subsystems. The states of each subsystem also affect the states and agents of neighboring subsystems, hence is conceptually similar to cellular automata described below.

The actions of these agents of change can be modelled as *transformations* on the states of each subsystem. Let $\mathbf{x}'$ denote the resulting state from the operation of a transformation function T_i. Now T_i may, in effect, be a function of not just the state of entity i, *i.e.*, $\mathbf{x}_i$, but also the states of neighboring subsystems. Thus, $T_i(\mathbf{x}_i) = f(\ldots \mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1} \ldots)$ where the states of neighboring subsystems are explicitly denoted and f is a vector-valued function. Subsequent states, those resulting from the action of the agents of change can be related to the current states by

$$\mathbf{x}'_i = f(\ldots \mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1} \ldots) = \hat{f}(\mathbf{x}_i)$$

for some $\hat{f} \in \hat{F}_i$ that is functionally dependent on the states of the neighbors of i. Because of our first and second formalisms, these transformations must restrict the output states to the set of Pareto optima P or near Pareto optima P', *i.e.*, $\mathbf{x}' \in S = P \cup P'$. In effect, the neighboring subsystems select transformations from a family $\hat{F}_i$ of transformations all of which map one point in S onto another point in S. Assuming the existence of identity and inverse transformations for each such point, the set $\hat{F}_i$ constitutes a *symmetric group* ([19]) over the states S.

2) Cellular Automata: Complexity theory, in particular theories regarding *cellular automata* (CA) share a number of features with SI (see [6, p. 245]). CAs are characterized by a set of *cells* usually arranged in some geometric pattern (although this is not technically required) and connected to other similar cells by some neighborhood rule. Each cell has a finite number of states which are determined by the states of its immediate neighbors. State transitions are defined by simple rules, a central component of cellular automata, that produce complex patterns often described as self-organized [36], [3]. The distinctions between SI and CA as proposed here involve the connection between the emergent behavior and Pareto optimality. CAs usually do not involve objective functions *per se*. It is also worth noting that there is a deep connection between CA and group theory [3, Ch. 3: Group CA Characterization]. Chaudhuri, *et al.* [3] provide a number of theoretical results on the connection between CA and cyclic groups and other group properties.

It is also worth noting that the transformation functions that determine how each cell evolves are usually very simple functions—somewhat analogous to the limited forms of communication associated with SI and stigmergy. But in SI, the relationships among the entities is dynamic owing to the mobility of the agents of change. Entity relationships are not fixed as in CAs. Notwithstanding this distinction, SFSMs described above can model this mobility in a way consistent with a CA paradigm by incorporating system states that model changing environmental conditions.

3) Multi-Objective Optimization: One important aspect in making a connection between self-organization and Pareto optima is how these Pareto optima are identified. By creatively utilizing the Lebesgue measure, SA and other heuristics can be engineered to search for these operating points in a very direct way, by maximizing this Lebesgue measure. This can be done using the COSA concept which happens to incorporate simple signalling systems as well. Many useful analogies to SI thus become apparent. Swarms of insects are analogous to parallel processors, simple signals in SI are analogous to the scalar Lebesgue measure or some estimate of it, maximization of this Lebesgue measure is analogous to movement along a Pareto optimal frontier, hence a type of self-organized behavior. This also suggests an interesting, and perhaps profound, unity of concepts between and among the various paradigms described here.

4) Game Theory: The dynamics of insect colonies involve competition and cooperation among them and so unavoidably involve aspects of game theory. The dynamics of competition and cooperation constitute the major components of the more interesting aspects of game theory [40]. Game theoretic concepts such as the famous *Nash Equilibrium* are really statements about Pareto optimality and how interacting entities achieve the greatest utility in a changing environment of competition and cooperation [41, p.193,339].

5) Markov Random Fields: The theory of *Markov Random Fields* (MRFs) shares a great deal in common with that of Boltzmann Machines, SA and CA. As alluded to earlier in the paragraphs on *Efficient Tradeoffs* . . .), some subsystem states may not be allowed to coexist with neighboring subsystem states. MRFs can model this type of behavior and have the desirable Markov property in terms of spatial attributes. This provides a mathematical structure to these problems. A large body of useful theoretical results on MRFs exists (se *e.g.,* [42]) and may be quite useful in addressing some of the issues in SI and how efficient tradeoffs of efficiency measures may be analyzed.

IV. CONCLUSION

The growth of communications and networked systems imposes on a us a need to explore new and imaginative ways to address the expected problems in managing these systems. Effectively utilizing the SI paradigm requires some frameworks on which to first build the relevant theory and then guide research toward practical implementations. This article examined three aspects of a meta-formalism that attempt to provide this necessary structure to further research into SI and

develop a more rigorous and mathematically sound theory for its analysis. The three elements of the meta-formalism involve 1) a set of first principles based on the laws of nature, 2) a dynamical framework, and 3) a problem framework. The relevant laws of nature are those associated with the theory of evolution and the effects of natural selection. The theory of evolution provides the justification for describing efficient behavior in social insect colonies. This efficient behavior can be mathematically characterized and associated with Pareto optima. Thus, the self-organized behavior often associated with SI can be given a more mathematical treatment when described in terms of operating points that lie along a Pareto optimal frontier. Operation along this frontier also provides a way of mathematically describing adaptive behavior, one of the hallmarks of SI. These notions of efficiency and adaptability which are implications of the first principles can be given effect as the underlying concepts behind the second formalism component—the dynamical framework.

The dynamical framework was described in terms of scale invariant Pareto optimality (SIPO), a property that requires that the behavior of entities within a swarm must be consistent with efficient behavior and tradeoffs of efficiency on many scales. SIPO thus provides a valuable set of conceptual constraints that focusses the research efforts, provides ideas that can be mathematically formalized and expressed, and features some of the mathematical elegance of *self-similarity* so often found in natural systems.

Finally, the problem framework provides a recipe for abstracting the problem and, in effect, provides the modelling clay with which to mold SI research so that it is consistent with the first two formalism components. The problem framework employs the use of swarming finite state machines models which provide a way to model a continuum of complexity apparent in the SI exhibited by different species.

In conclusion, it is the author's hope that this meta-formalism and the ideas in this article will help facilitate research and development using SI. The resulting theories have potential for application in many areas of inquiry involving complex systems. The goal of making these increasingly complex systems more efficient, scalable and autonomous will help to ensure their effectiveness well into the future.

ACKNOWLEDGEMENT

DISCLAIMER

The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory, the National Aeronautics and Space Administration, or the U.S. Government.

REFERENCES

[1] A. Leon-Garcia and I. Widjaja, *Communication Networks: Fundamental Concepts and Key Architectures*, ser. Computer Science and Electrical and Computer Engineering.   New York, NY: The McGraw-Hill Companies, Inc., 2000.

[2] *Making the Nation Safer: The Role of Science and Technology in Countering Terrorism*.   Washington, D.C.: National Academy Press, 2002, ch. 5: Information Technology, pp. 135–176, a Report to the National Academies from the Committee on Science and Technology for Countering Terrorism.

[3] P. Chaudhuri, D. Chowdhury, S. Nandi, and S. Chattopadhyay, *Additive Cellular Automata: Theory and Applications*.   New York, NY: John Wiley & Sons, Inc., June 1997, vol. 1.

[4] W. Poundstone, *The Recursive Universe: Cosmic Complexity and the Limits of Scientific Knowledge*.   Chicago, Ill.: Contemporary Books, 1985.

[5] I. Prigogine and I. Stengers, *Order Out of Chaos: Man's New Dialog with Nature*.   New York, NY: Bantam Books, Inc., 1984.

[6] E. Bonabeau, M. Dorigo, and G. Theraulaz, *Swarm Intelligence: From Natural to Artificial Systems*, ser. Santa Fe Institute Studies in the Sciences of Complexity.   New York, N.Y.: Oxford University Press, 1999.

[7] B. Bullnheimer, R. F. Hartl, and C. Strauss, "An improved ant system algorithm for the vehicle routing problem," University of Vienna, Vienna, Austria, Tech. Rep., POM Working Paper No. 10/97, 1997.

[8] V. Maniezzo, A. Colorni, and M. Dorigo, "The ant system applied to the quadratic assignment problem," Universite Libre de Bruxelles, Belgium, Tech. Rep., Tech. Rep. IRIDIA/94-28 1994.

[9] R. Schoonderwoerd, O. Holland, J. Bruten, and L. Rothkrantz, "Ant-based load balancing in telecommunications networks," *Adapt. Behav.*, vol. 5, pp. 169–207, 1996, early work on SI in networks.

[10] L. Kaelbling, M. Littmand, and A. Moore, "Reinforcement learning: A survey," *Journal of Artificial Intelligence Research*, vol. 4, pp. 237–285, 1996.

[11] M. Littman and J. Boyan, "A distributed reinforcement learning scheme for network routing," Carnegie Mellon University, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, Tech. Rep., 1993.

[12] G. Di Caro and M. Dorigo, "Antnet: A mobile agents approach to adaptive routing," Universite Libre de Bruxelles, Belgium, Tech. Rep., Tech. Report IRIDIA/97-12, 1997.

[13] H. V. D. Parunak, M. Purcell, and R. O'Connell, "Digital pheromones for autonomous coordination of swarming uav's," American Institute of Aeronautics and Astronautics, Tech. Rep. AIAA 2002-3446, 2002.

[14] D. Subramanian, P. Druschel, and J. Chen, "Ant and reinforcement learning: A case study in routing in dynamic networks," Rice University, Department of Computer Science, Tech. Rep.

[15] M. Millonas, "Artificial life iii," ser. Sante Fe Institute Studies in the Sciences of Complexity, C. Langton, Ed., vol. XVII.   New York, NY: Addison-Wesley Publishing Co., 1994, pp. 417–445.

[16] N. Wiener, *Cybernetics: Or Control and Communication in the Animal and the Machine*, 2nd ed.   Cambridge, MA.: The M.I.T. Press, Inc., 1961.

[17] G. Di Caro and M. Dorigo, "Antnet: Distributed stigmergetic control for communications networks," *Journal of Artificial Intelligence Research*, vol. 9, pp. 317–365, 1998.

[18] R. Serra and G. Zanarini, *Complex Systems and Cognitive Processes*. New York, NY: Springer-Verlag, 1990.

[19] J. Rosen, *Symmetry in Science: An Introduction to the General Theory*. New York, NY: Springer-Verlag, 1995.

[20] E. Aarts and J. Korst, *Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing*.   John Wiley & Sons, 1989.

[21] I. Bohachevsky, M. Johnson, and M. Stein, "Generalized simulated annealing for function optimization," *Technometrics*, vol. 28, no. 3, pp. 209–217, 1986.

[22] F. R. Mitra, D. and A. Sangiovanni-Vincentelli, "Convergence and finite-time behavior of simulated annealing," *Advances in Applied Probability*, vol. 18, pp. 747–771, 1986.

[23] M. A. Fleischer and S. H. Jacobson, "Scale invariance properties in the simulated annealing algorithm," *Methodology and Computing in Applied Probability*, vol. 4, pp. 219–241, 2003.

[24] N. A. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *J. of Chemical Physics*, vol. 21, pp. 1087–1092, 1953.

13

[25] S. Kirkpatrick, C. D. G. Jr., and M. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, 1983.

[26] J. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor, Michigan: University of Michigan Press, 1975.

[27] C. Reeves, *Modern Heuristic Techniques for Combinatorial Problems*. New York, NY: John Wiley & Sons, Inc., 1993.

[28] M. A. Fleischer, "The measure of pareto optima: Applications to multiobjective metaheuristics," ser. Proceedings of the Second International Conference on Evolutionary Multi-Criteria Optimization—EMO 2003, vol. Lecture Notes in Computer Science. Faro, Portugal: Springer-Verlag, 2002, p. To appear.

[29] ——, "The measure of pareto optima: Applications to multiobjective metaheuristics," Institute for Systems Research, University of Maryland, College Park, MD., Tech. Rep. 2002-32, 2002.

[30] C. Coello and M. Lechuga, "Mopso: A proposal for multiple objective particle swarm optimization," Evolutionary Computation Group at CINVESTAV, CINVESTAV-IPN, Mexico, Technical Report EVOCIV-01-2001, September 2001.

[31] F. Menczer, M. Degeratu, and W. Street, "Efficient and scalable pareto optimization by evolutionary local selection algorithms," *Evolutionary Computation*, vol. 8, no. 2, pp. 223–247, 2000.

[32] E. Zitzler and L. Thiele, "Multiobjective optimization using evolutionary algorithms– a comparative case study," in *Proceedings of the Fifth International Conference on Parallel Problem Solving from Nature—PPSN V*, M. S. H. S. A.E. Eiben, T. Bäck, Ed., Berlin, Germany, 1998.

[33] R. Dorf and R. H. Bishop, *Modern Control Systems*, 9th ed. Upper Saddle River, NJ: Prentice-Hall, Inc., 2001.

[34] M. A. Fleischer, "Cybernetic optimization by simulated annealing: Accelerating convergence by parallel processing and probabilistic feedback control," *Journal of Heuristics*, vol. 1, no. 2, 1996.

[35] ——, *Metaheuristics: Advances and Trends in Local Search Paradigms for Optimization*. Kluwer Academic Publishers, 1999, ch. 28: Generalized Cybernetic Optimization: Solving Continuous Va riable Problems, pp. 403–418.

[36] Y. Bar-Yam, *Dynamics of Complex Systems*. Reading, MA.: Perseus Books, 1997.

[37] A. Smith, *An Inquiry Into The Nature and Causes of The Wealth of Nations*, ser. Everyman's Library. J.M. Dent & Sons Ltd., 1910, vol. I.

[38] A. Toffler, *The Third Wave*. New York, NY: Bantam Books, 1991.

[39] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA: The MIT Press, 1998.

[40] T. C. Schelling, *The Strategy of Conflict*. New York, NY: Oxford University Press, 1960,77, a most excellent text on the science of conflict.

[41] R. D. Luce and H. Raiffa, *Games and Decisions: Introduction and Critical Survey*. New York, NY.: Dover Publications Inc., 1985.

[42] P. Bremaud, *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. New York, NY: Springer Verlag, 1999.

Dr. Mark Fleischer is an Assistant Research Scientist in the Center for Satellite and Hybrid Communications Networks at the Institute for Systems Research at the University of Maryland, College Park. He has a varied and multi-disciplinary background: a bachelor's degree in political science from M.I.T., a law degree from Cleveland State University, and a Ph.D. in operations research from Case Western Reserve University. His interests span a number of areas involving dynamical systems, stochastic processes, applied probability, computer networking, and control theory. His recent research involves optimization techniques in the domain of multi-objective optimization where he developed results concerning the measure of Pareto optima. His prior research involved discerning the connections between the simulated annealing optimization technique and information theory. He has also developed new forms of parallelization techniques for simulated annealing by using feedback control mechanisms to modulate the algorithm's behavior. A recent paper of his describes his discovery of scale invariant properties in the simulated annealing algorithm and will soon appear in an upcoming journal article. These optimization techniques have also been used in conjunction with new and general resource allocation and scheduling models and he continues to work on systems analysis, integration and decision problems particularly in the area of swarm intelligence.